Klasické metody redukce dimenze dat

1. Připomenutí PCA (Principal Component Analysis, rozklad na hlavní komponenty)
2. ICA (Independent Component Analysis, rozklad na nezávislé komponenty) [Jutten & Hérault 1991]
3. Fisherova lineární diskriminační analýza (FLDA)

Literatura k tématu

[1] R.O. Duda – P.E. Hart – D.G. Stork. Pattern Classification. John Wiley & Sons, 2001.

[1] A. Hyvärinen – E. Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5):411-430, 2000.

PCA: Rozklad na hlavní komponenty

Máme množinu bodů $\mathbf{x}_1, \ldots, \mathbf{x}_k$ v prostoru dimenze d, z bodů vytvoříme $d \times k$ matici $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$

Hledáme nejlepší reprezentaci $\mathbf{X}$ v prostoru dimenze $s < d$

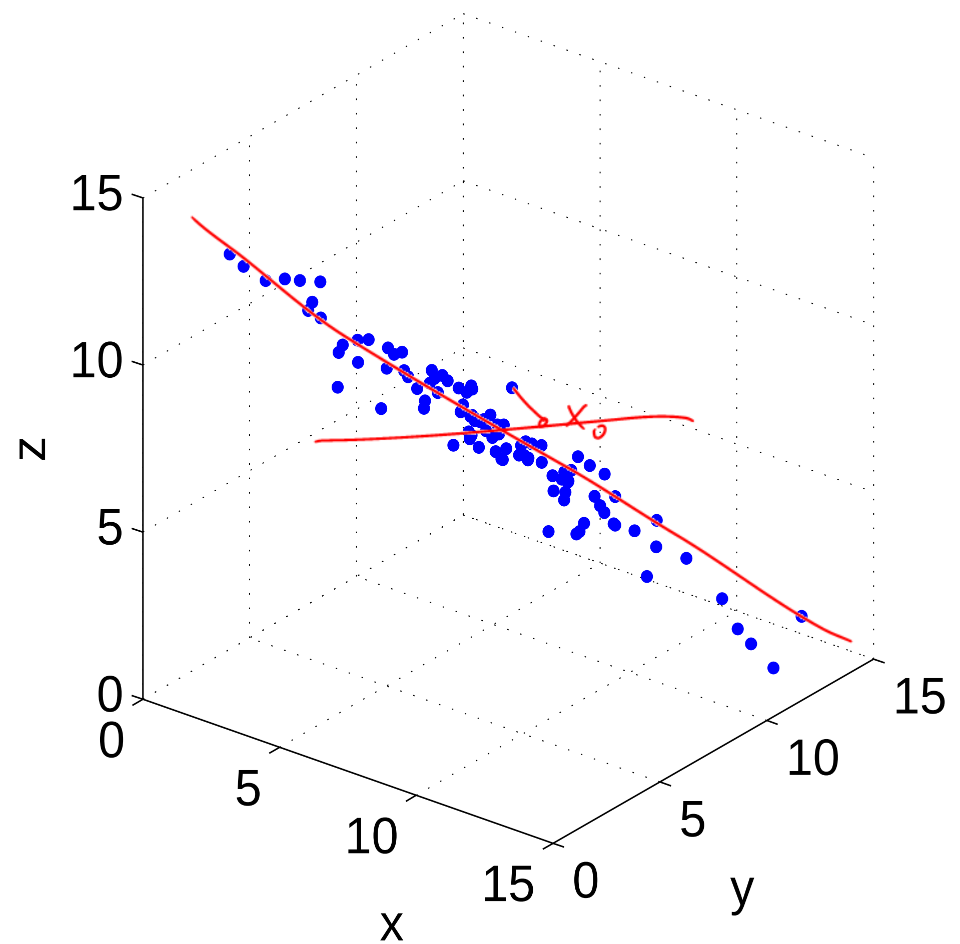
$X = (x_1, y_2, x_3)$

1. Nejlepší reprezentace pro $s = 0$ minimalizuje energii

$$J_0(\mathbf{x}_0) = \sum_{i=1}^{k} \|\mathbf{x}_i - \mathbf{x}_0\|^2,$$

kde řešením je

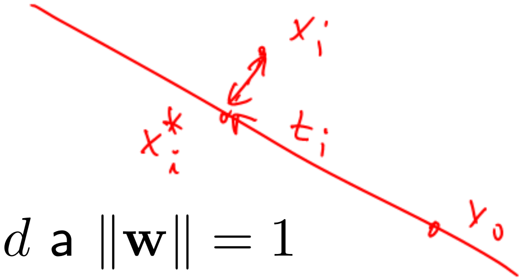
$$\mathbf{x}_0 = \frac{1}{k} \sum_{i=1}^{k} \mathbf{x}_i \qquad \text{důkaz: } \circledast 1$$



$d \times k \quad k \times d$

- pozn: $\mathbf{X}\mathbf{X}^\top$ je $d \times d$ kovarianční matice $\sum_{i=1}^{k} \mathbf{x}_i \mathbf{x}_i^\top$

PCA: pokračování



2. Nejlepší reprezentace při $s = 1$ je průmět na přímku

$$\mathbf{y}_i = \mathbf{x}_0 + t_i \mathbf{w}, \quad \text{kde } \mathbf{w} \text{ je vektor dimenze } d \text{ a } \|\mathbf{w}\| = 1$$

neznámé

Parametry t_i, $i = 1, 2, \ldots, k$ a $\mathbf{w}$ minimalizují energii

$x_i - x_0 = \bar{x}_i$

$$J(t_1, \ldots, t_k, \mathbf{w}) = \sum_{i=1}^{k} \|\underbrace{\mathbf{x}_0 + t_i \mathbf{w}}_{y_i} - \mathbf{x}_i\|^2 = \sum_i \|t_i \mathbf{w} - \bar{\mathbf{x}}_i\|^2 = \sum_i (t_i \mathbf{w} - \bar{\mathbf{x}}_i)^\top (t_i \mathbf{w} - \bar{\mathbf{x}}_i) =$$

$$= \sum_{i=1}^{k} \left(t_i^2 - 2 t_i\, \mathbf{w}^\top \bar{\mathbf{x}}_i + \bar{\mathbf{x}}_i^\top \bar{\mathbf{x}}_i\right)$$

$2t_i - 2\, w^\top \bar{x}_i = 0$

z $\frac{\partial J}{\partial t_i} = 0$ dostaneme $t_i = \mathbf{w}^\top \bar{\mathbf{x}}_i = \bar{\mathbf{x}}_i^\top \mathbf{w}$. Po dosazení

$Q = w^\top S w + \lambda (w^\top w - 1)$

$- w^\top \bar{x}_i \bar{x}_i^\top w \quad + \bar{x}_i^\top \bar{x}_i$

nutná: $Sw = \lambda w$

$$J(t_1, \ldots, t_k, \mathbf{w}) = \cdots = \sum_{i=1}^{k} \bar{\mathbf{x}}_i^\top \bar{\mathbf{x}}_i - \mathbf{w}^\top \left(\underbrace{\sum_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^\top}_{XX^\top} \right) \mathbf{w} = \sigma^2 - \mathbf{w}^\top \mathbf{S}\, \mathbf{w},$$

$\underset{\|w\|^2 = 1}{\operatorname{argmax}}\ w^\top S w$

kde $\mathbf{S}$ je $d \times d$ kovarianční matice $\mathbf{S} = \mathbf{X}\mathbf{X}^\top$

PCA: pokračování

Hledáme minimum J za podmínky $\|\mathbf{w}\|^2 = 1$

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \left(\sigma^2 - \mathbf{w}^\top \mathbf{S}\mathbf{w} + \lambda(\mathbf{w}^\top\mathbf{w} - 1)\right)$$

z $\frac{\partial J}{\partial \mathbf{w}} = 0$ dostaneme $2\mathbf{S}\mathbf{w} - 2\lambda\mathbf{w} = 0$, z čehož plyne

> $\mathbf{w}$ musí být vlastním vektorem $\mathbf{S} = \mathbf{X}\mathbf{X}^\top$ odpovídajícím největšímu vlastnímu číslu.

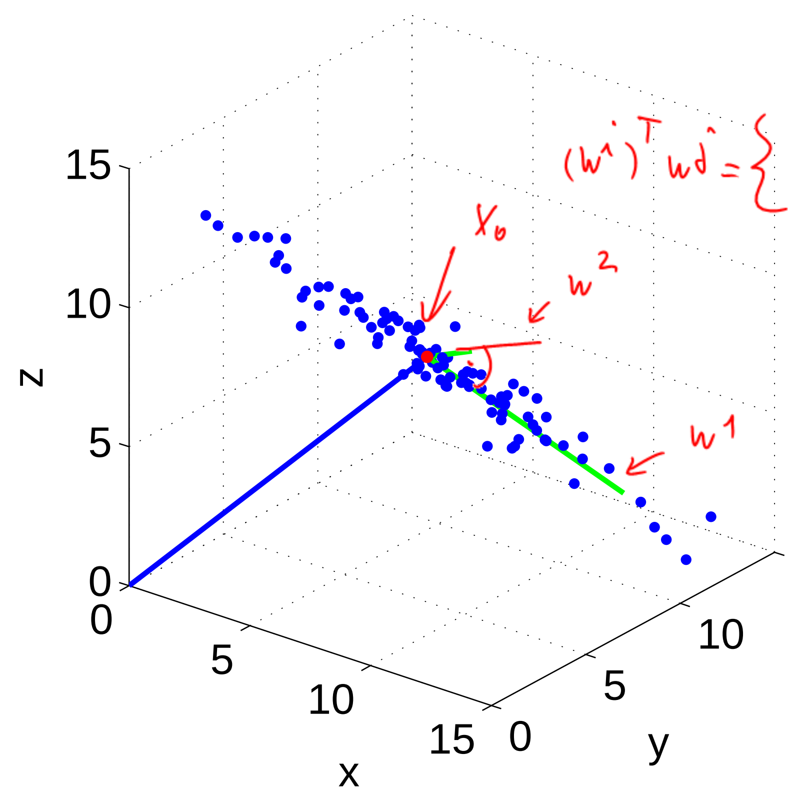
3. při $s > 1$ postupujeme obdobně, promítáme do s-dimenzionální nadroviny

$$\mathbf{y}_i = \mathbf{x}_0 + \sum_{j=1}^{s} t_i^j \mathbf{w}^j$$

$s < d$

$(w^i)^\top w^j = \begin{cases} 1 & i = j \\ 0 & \end{cases}$

a zjistíme, že $\mathbf{w}^j$ jsou rovny **ortogonálním vlastním vektorům** $\mathbf{S}$ odpovídající s největším vlastním číslům.



Vlastnosti PCA

- vybíráme komponenty v pořadí energie (rozptylu)
- komponenty jsou nekorelované (po odečtení střední hodnoty $\mathbf{x}_0$):

$$(t_i^m \mathbf{w}^m)^\top (t_i^n \mathbf{w}^n) = t_i^m \, t_i^n (\mathbf{w}^m)^\top \mathbf{w}^n = \begin{cases} t_i^{2n} & \text{pro } m = n \\ 0 & \text{jinak} \end{cases}$$

- aby metoda měla smysl, kovarianční matice má mít plnou hodnost, tedy $k > d$
- pokud nás zajímá ‚signál' promítnutý do s hlavních komponent PCA, pak

s×k + d×s   d×k

$$\mathbf{y} = \mathbf{W}^\top (\mathbf{x} - \mathbf{x}_0),$$

d×s

kde $\mathbf{W}$ je $d \times s$ matice hlavních komponent $\mathbf{W} = [\mathbf{w}^1, \mathbf{w}^2, \ldots, \mathbf{w}^s]$ po sloupcích

- komponenty $\mathbf{w}^j$ jsou „vzorové signály“, lineární filtry s velkou odezvou na data podobná předloženým interpretace viz příklad na str. 7

Rekonstrukce signálu z PCA

- rekonstrukce původního signálu je, po jednotlivých vektorech $\mathbf{x}_j^*$

$$\mathbf{x}_j^* = \sum_{m=1}^{s} \mathbf{w}^m \alpha_m, \quad \text{kde} \quad \alpha_m = (\mathbf{w}^m)^\top (\mathbf{x}_j - \mathbf{x}_0)$$

Karhunen-Loève

- po přepsání

$$\mathbf{x}_j^* = \sum_{m=1}^{s} \mathbf{w}^m (\mathbf{w}^m)^\top (\mathbf{x}_j - \mathbf{x}_0) = \mathbf{P}(\mathbf{x}_j - \mathbf{x}_0),$$

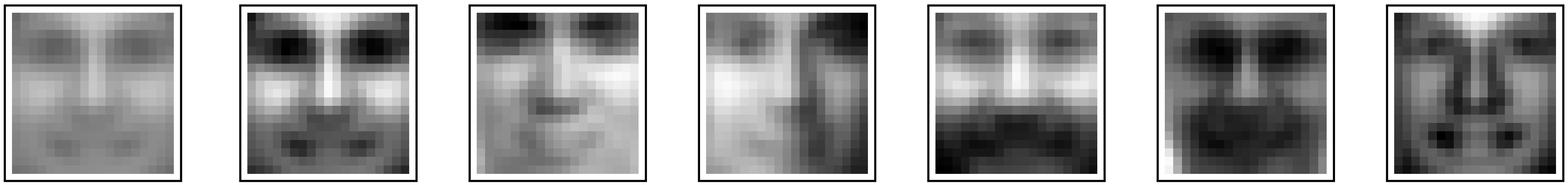
- $\mathbf{P} = \mathbf{W}\mathbf{W}^\top$ je singulární $d \times d$ matice hodnosti s
- nulový prostor $\mathbf{P}$ odpovídá potlačeným složkám signálu
- $\mathbf{P}$ je tzv. ‚projekční matice‘ : $\mathbf{P}^2 = \mathbf{P}$
- rekonstrukce $\mathbf{P}(\mathbf{x}_j - \mathbf{x}_0) + \mathbf{x}_0$ je ‚filtrovaný signál‘
- PCA minimalizuje energii reziudua $\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_0)$ ⊛1

Aplikace: Eigenfaces

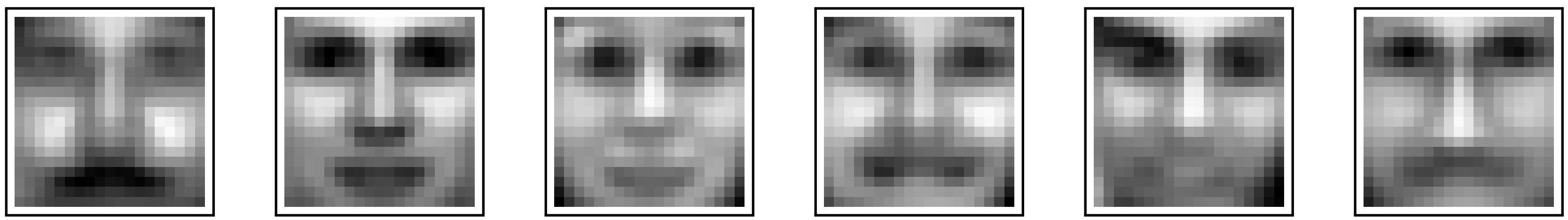
x_i

2901 snímků 19×19 pixelů: data laskavostí MIT http://www.ai.mit.edu/projects/cbcl

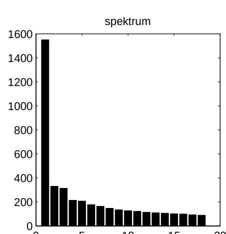
 . . .

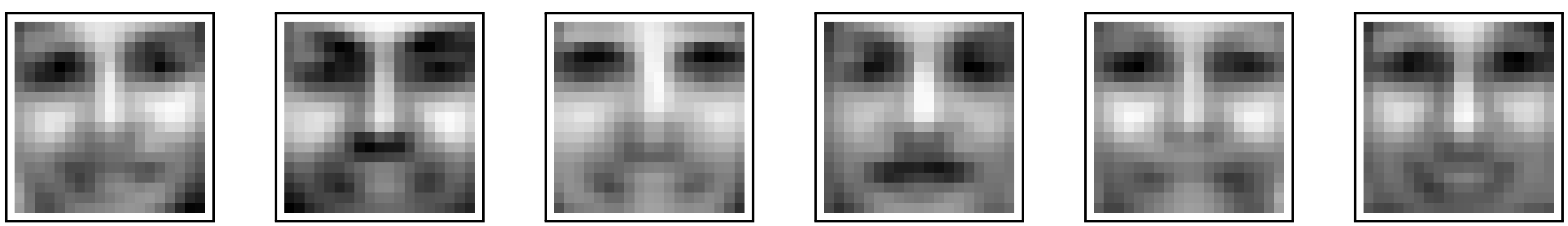


$\mathbf{x}_0$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^1$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^2$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^3$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^4$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^5$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^6$



$\mathbf{x}_0 + \alpha\mathbf{w}^7$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^8$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^9$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{10}$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{11}$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{12}$





$\mathbf{x}_0 + \alpha\mathbf{w}^{13}$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{14}$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{15}$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{16}$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{17}$ $\quad$ $\mathbf{x}_0 + \alpha\mathbf{w}^{18}$

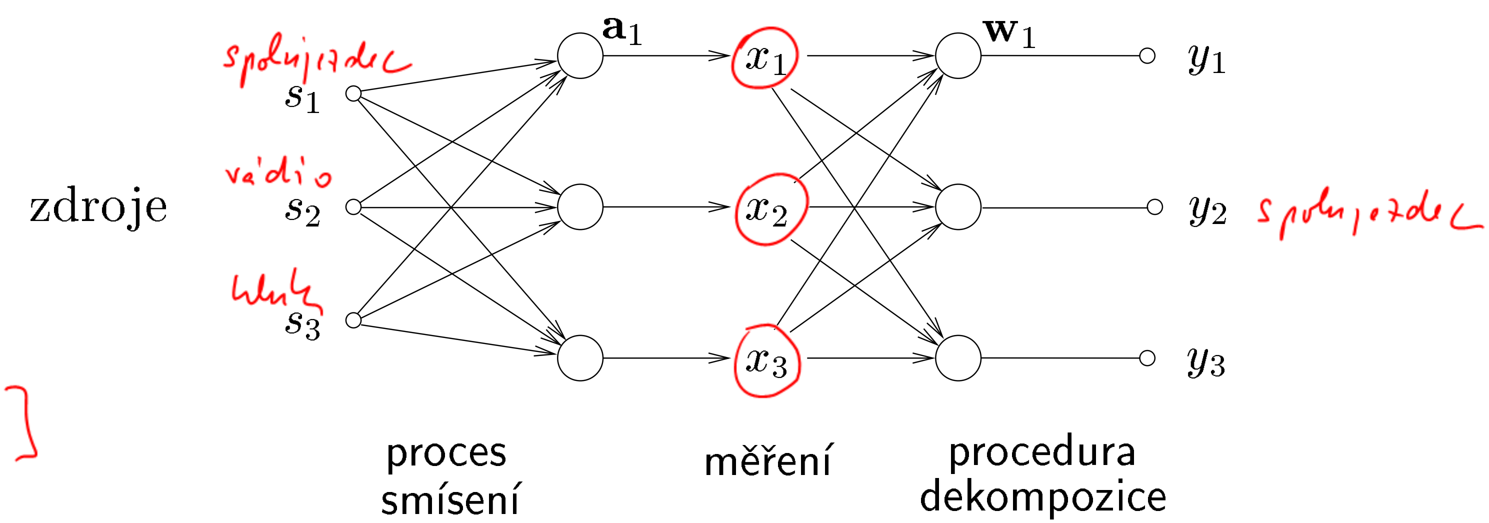
- průměty do vybraných komponent jsou příznaky k rozpoznání tváří od ‚netváří'
- pro klasifikátor nutný (ruční) výběr komponent, které nemodelují osvětlení
- redukce z dimenze 361 do nižší potlačí šum; rekonstrukce prvního příkladu z 18 komponent:

$P(x - y_0) + y_0$

- dobrý klasifikátor vyžaduje též PCA všech ‚netváří'

Rozklad na nezávislé komponenty

Předpokládáme, že měření x_i je neznámou směsí signálu z malého (známého) počtu neznámých nezávislých zdrojů $\mathbf{s} = [\,s_1, \ldots, s_s\,]^\top$, tj. $x_i = \mathbf{a}_i^\top \mathbf{s}$



$X = [x_1, \ldots, x_k]$

dim 3

- hledáme komponenty $y_i = \mathbf{w}_i^\top \mathbf{X}$ tak, aby měly vlastnosti původních složek s_i

$$\mathbf{y} = \mathbf{W}^\top \mathbf{X} = \mathbf{W}^\top \mathbf{A}^\top \mathbf{s}$$

$\mathbf{W}$ bude ‚inverzí‘ $\mathbf{A}$

- nebude možné určit rozptyl $\mathbf{s}_i$, protože $x_i = \mathbf{a}_i^\top \mathbf{s} = \mathbf{a}_i^\top \sigma \frac{1}{\sigma} \mathbf{s}$
- $\Rightarrow$ komponenty nebudou mít ‚pořadí‘ jako v PCA
- Jak poznat s? Nezávislost $\mathbf{s} \Rightarrow$ nezávislost $\mathbf{y}$
- **Postup:** hledání $\mathbf{W}$ tak, aby $\mathbf{y}$ byly nezávislé

Rozklad na nezávislé komponenty: ICA

Co můžeme dostat?

1. pomocí PCA dostaneme nekorelované složky, ale ne nezávislé, protože když místo $x_i = \sum_j a_{ij} s_j$ bude $x_i = \sum_j a_{ij} f(s_j)$, tak dostaneme jiné řešení
2. zdroje musí být negausovské náhodné proměnné, protože ortogonální transformace $\mathbf{R}$ gaussovské náhodné proměnné je gaussovská a u normálního rozdělení je nekorelovanost a nezávislost totéž $x_i = \mathbf{a}_i^\top \mathbf{s} = \mathbf{a}_i^\top \mathbf{R}^\top \mathbf{R}\, \mathbf{s} \Rightarrow$ ‚nepoznáme své komponenty‘

- $\Rightarrow$ nelze použít PCA, protože předpokládá normalitu
- negausovské rozdělení má vyšší momenty než do 2. řádu

ICA: míra nezávislosti

Hledáme $y_i = \mathbf{w}_i^\top \mathbf{X}$ tak, aby y_i byly navzájem nezávislé, tj, aby platilo

$$p(y_1, \dots, y_s) = \prod_{i=1}^{s} p(y_i)$$

Pokud jsou y_i nezávislé, platí

tento vzorec pro $h_i(p) = \log(p)$ známe z vlastností entropie

$$E\{h_1(y_1) \cdot h_2(y_2) \cdots h_s(y_s)\} = \prod_{i=1}^{s} E\{h_i(y_i)\}$$

Vhodnou mírou nezávislosti je vzájemná informace

$$I(y_1, \dots, y_s) = \sum_{i=1}^{s} H(y_i) - H(y_1, \dots, y_s)$$

pokračování

w^i

Problém $y_i = \mathbf{w}_i^\top \mathbf{X}$, $i = 1, 2, \ldots, s$ lze zapsat v maticovém tvaru $\mathbf{y} = \mathbf{W}^\top \mathbf{X}$, kde $\mathbf{W}$ je matice $s \times d$

Pokud $\mathbf{y} = \mathbf{W}^\top \mathbf{X}$, tak platí

$w_i^\top X$

$$I(y_1, \ldots, y_s) = \sum_i H(y_i) - \cancel{H(x_1, \ldots, x_k)} - \log \det \mathbf{W}$$

protože

$y_i = w_i^\top X$

$$H(y_1, \ldots, y_s) = \iint f(y_1, \ldots, y_s) \log f(y_1, \ldots, y_s)\, \mathrm{d}y_1 \cdots y_s = \iint f(\mathbf{x}) \log \frac{f(\mathbf{x})}{\det \mathbf{W}}\, \mathrm{d}\mathbf{x}$$

Když y_i budou mít jednotkovou kovarianční matici, tak

$$1 = \det \mathbf{I} = \det \mathbf{y}\mathbf{y}^\top = \det \left(\mathbf{W}^\top \mathbf{X}\mathbf{X}^\top \mathbf{W} \right) = (\det \mathbf{W})^2 \det \mathbf{X}\mathbf{X}^\top$$

takže $\det \mathbf{W}$ je konstanta. Z toho plyne, že

$$I(y_1, \ldots, y_s) = C + \sum_{i=1}^{s} H(y_i)$$

ICA: aproximace kriteriální funkce

Hledáme tedy $\mathbf{w}_i$ tak, že

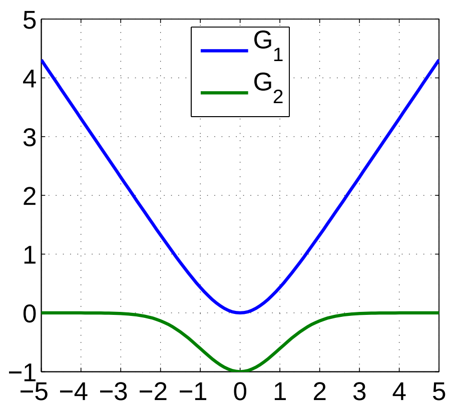
$$\mathbf{W}^* = \arg \overset{\text{min}}{\cancel{\max}}_{\mathbf{W}} \sum_{i=1}^{s} H(\mathbf{w}_i^\top \mathbf{X}), \qquad \text{kde } \mathbf{w}_i \text{ jsou řádky matice } \mathbf{W}^\top$$

Z výpočetních důvodů se entropie $H(u) = E\{-\log(u)\}$ aproximuje funkcí

$$J(y) = \big(E\{G(y)\} - E\{G(\nu)\}\big)^2,$$

kde $\nu \sim N(0, \mathbf{I})$ a typické volby jsou

$$G_1(u) = \frac{1}{a_1} \log\cosh a_1 u \qquad \text{nebo} \qquad G_2(u) = -e^{-a_2 \frac{u^2}{2}}$$



$a_1 = a_2 = 1$

$G(u) = u^2$ nelze použít, protože ‚nerozpozná' nezávislost od nekorelovanosti

Procedura ICA

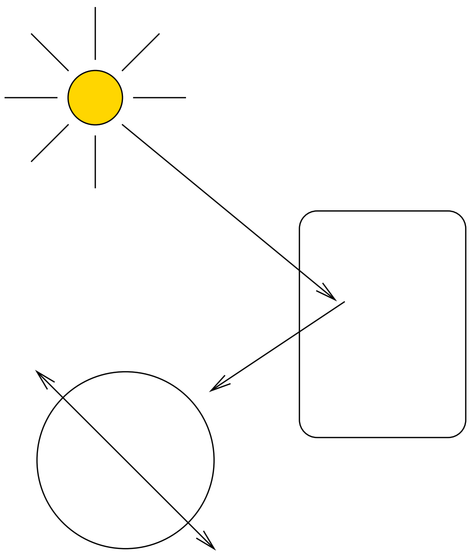
Postup pro $s = 1$

1. normalizuj data $\bar{\mathbf{x}}_i := \mathbf{S}^{-\frac{1}{2}}(\mathbf{x}_i - \mathbf{x}_0)$ pro $i = 1, 2, \ldots, k$, „bělení"
   kde $\mathbf{S} = \sum_{i=1}^{k} (\mathbf{x}_i - \mathbf{x}_0)(\mathbf{x}_i - \mathbf{x}_0)^\top$ je kovarianční matice a $\mathbf{S} = \mathbf{S}^{\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}$ Choleski
2. zvol ~~náhodný~~ libovolný vektor $\mathbf{w}$ tak, aby $\|\mathbf{w}\| = 1$
3. nechť $\mathbf{w}^+ = \sum_{i=1}^{k} \bar{\mathbf{x}}_i \, G'(\mathbf{w}^\top \bar{\mathbf{x}}_i) - G''(\mathbf{w}^\top \bar{\mathbf{x}}_i)\,\mathbf{w}$ Newtonovská iterace
4. renormalizuj $\mathbf{w} = \dfrac{\mathbf{w}^+}{\|\mathbf{w}^+\|}$
5. opakuj kroky 3–4 do konvergence
6. nalezená komponenta je $y = \mathbf{w}^\top \bar{\mathbf{X}}$ y je vektor délky k

odvození nutné podmínky pro minimum (proměnná je $\mathbf{w}$)
nutná podmínka pro $J(\mathbf{w}^\top \mathbf{x}) \longrightarrow$ nutná podmínka pro $E\{G(\mathbf{w}^\top \mathbf{x})\}$ za podmínky $\|\mathbf{w}\|^2 = 1$.

- dobrá implementace je podstatně složitější

Aplikace ICA: separace odlesků



analyzátor
polarizace

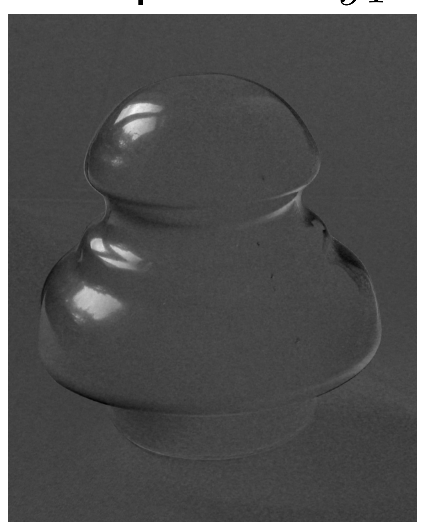
vstupní obrazy

x_1

x_2

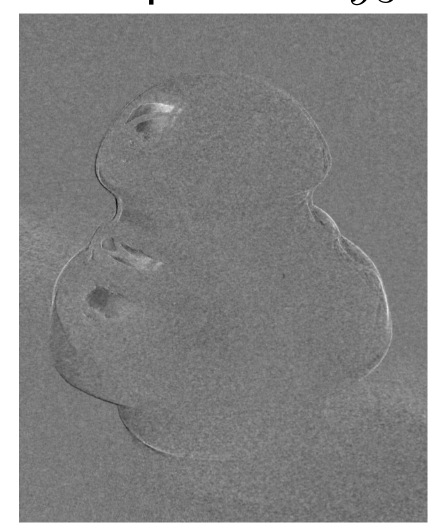
x_3

komponenta y_1



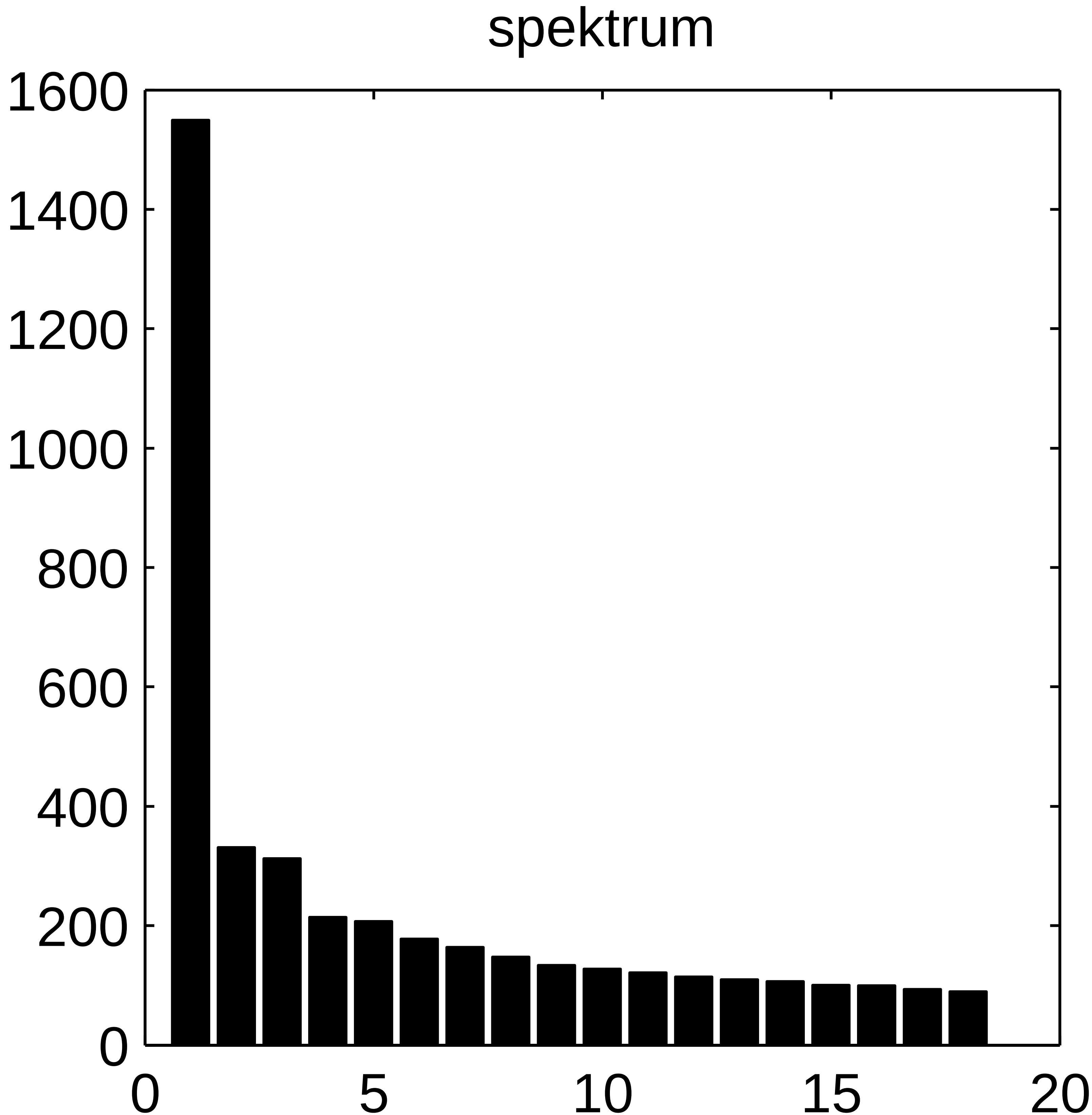
komponenta y_2

komponenta y_3



- rovina polarizace závisí na úhlu normály $\Rightarrow$ pohyb odlesku
- 3. komponenta částečně obsahuje tento pohyb

Konec


spektrum
1600
1400
1200
1000
800
600
400
200
0
0
5
10
15
20